TRIPOD OPERATORS FOR REALTIME RECOGNITION OF SURFACE SHAPES IN RANGE IMAGES

Frank Pipitone
Navy Center for Artificial Intelligence,
Naval Research Laboratory,
Washington, DC 20375-5337

ABSTRACT

Tripod operators (TO's) are a versatile class of feature extraction operators for surfaces. They are useful for recognition and/or localization based on range or tactile data. They extract a few sparse point samples in a regimented way, so that N sampled surface points yield only N–3 independent scalar features containing all the pose-invariant surface shape information in these points and no other information. They provide a powerful index into sets of prestored surface representations. A TO consists of three points in 3-space fixed at the vertices of an equilateral triangle and a procedure for making several "depth" measurements in the coordinate frame of the triangle, which is placed on the surface like a surveyor's tripod. TO's can be imbedded in a vision system in many ways and applied to almost any surface shape. Here the focus is an experimental study in which individual TO's are used to search a cluttered range image for one of 25 known shapes, typically in milliseconds, with very few false detections. We believe that this simple way of using TO's, in conjunction with existing triangulation range sensor technology, can be effectively applied to industrial parts recognition tasks, and with additional research, to other applications.

1. INTRODUCTION

This work is motivated by the long-standing observation that a small set (e.g., six to twelve) of point samples of the surface of an object is highly informative, and that it ought to be possible to construct a procedure for mapping such data into the identity and/or pose of an object in essentially constant time, for a significant range of cases. We have largely succeeded in doing this, using a geometric procedure called the tripod operator (TO). A typical TO is applied to a range image in approximately 2 milliseconds, as currently implemented on a Sun SPARCstation 10, resulting in a hypothesis about the surface under the operator. Software optimizations are expected to reduce this to well below 1 millisecond. A range image can be searched for a shape by repeatedly applying TO's at random places on the image. Potential applications include industrial parts recognition, target recognition, mobile robot vision, and face recognition.

In order to rapidly recognize objects based on surface shape, especially if the library of known objects is large and/or the average complexity of each object's surface shape is large, one needs to make feature measurements which are sufficiently informative, despite noise, that the reduction in the candidate set per unit computation time is acceptable. For example, one might reasonably measure this by the reduction in the Shannon Entropy of the set of identities and/or poses. By such a measure, steady progress has been made in previous work. Grimson [4,5] and others [6,7,8,9] extensively developed the idea of searching for associations between image features and model elements consistent with geometric constraints among the model elements, using *interpretation trees* to represent the consistent hypothesised associations (interpretations). However, interpretation trees require quadratic time processing per model. This is mitigated by using particularly informative features. We have argued that TO's can be used efficiently as such features [1,2]. A second connection is that a TO can be regarded as precompiled prepruned interpretation trees having sparse range pixels as the image features. This is their original inspiration. Lamdan and Wolfson [10] contribute to efficiency in model-based vision by providing precompiled geometric pointers among local features. This requires the ability to detect a reasonably small number of reasonably stable *interest points* and to define informative features there, whereas TO's are to be used anywhere on a surface and their informativeness can be looked up. Stein and Medioni [11] describe local operators called "splashes" with attractive invariance properties, but they have high

computational cost and depend on unoccluded and valid range pixels on certain geodesic lines. The RANSAC method [12] uses sparse samples economically to test a fit to a specific class of functions, but indexing is not provided; one must sequentially try function classes. Many kinds of local feature detectors or matchers have been explored for range images. For example, [13] concentrates on dihedral edges and [14] on the two principle curvatures of smooth surfaces. The principle limitation of most of them appears to be their discriminating power per unit computation. For example, estimators of the two principle curvatures either provide us with two real numbers worth of indexing information (and direction information), or the decision that the surface is not a good fit to a quadric in the current neighborhood. The former case allows discrimination of roughly q^2 local surface shapes if we can resolve q curvature values in noise. The latter case requires us to continue looking for local feature information (perhaps a dihedral will fit here or a quadric patch elsewhere). TO's provide one operation at a place on a surface, yielding a feature vector of any dimensionality d, applicable to nearly any surface, and potentially discriminating as many as roughly q^d hypotheses about the object (and/or its pose) on which the TO lies.

Tripod Operators are "somewhat global" and can sometimes straddle many surface undulations with its point samples, and span a large proportion of an object. They can operate on sparse regions of a dense range map, sparse data acquired actively from a sequential random access range scanner (such as in [15]), or via a tactile version of a TO. In earlier publications, we argue that the TO should allow very fast recognition [1] and present supporting experimental evidence by discriminating 1 object from a library of 10 using in some cases only *one* TO placement, using synthetic range data [2]. In [3] we extend this to the case of noisy LIDAR range images of isolated real objects, using a Bayesian approach to obtain reliable recognition using a small number (5 to 10) of low order (order 4) and high noise (1/10 the TO's edgelength) TO placements.

We have been studying TO's using a software system called TRIPOD, which allows various experiments to be performed involving the application of various kinds of TO's to real or synthetic range images, and the use of various representation and matching methods on the resulting feature space point sets. Our overall research goal is to determine the limits of performance of a vision system based on TO's, and to realize that performance in prototype vision systems. Performance measures of interest to us include speed, classification error, tolerance of noise and occlusion, library size, storage requirements, and ease of representing new shapes. Variables in such a vision system that effect performance include

1. Edge length of operator
2. Order of operator
3. Efficiency of the algorithm that computes the operator
4. What hypothesis verifier is used, if any
5. Representation of the TO invariant signatures
6. Indexing method used to assess proximity of TO measurement to signature
7. Method for relating multiple TO's on the same object
8. Method for representing pose constraints
9. Use of probabilistic reasoning

The focus of this paper is the use of isolated TO placements to rapidly recognize instances of a set of 25 typical manufactured surface shapes in range images containing a variety of known and unknown shapes. Items 4,7,8 and 9 above are outside the scope of this paper. Our two-part research strategy is to first learn how to obtain the greatest possible discrimination in the shortest time using individual TO placements, and in other work to exploit the relative pose of multiple placements to further increase performance.

2. REVIEW OF TRIPOD OPERATORS

2.1 Definitions and Properties
A tripod operator consists of three points in 3-space fixed at the vertices of an equilateral triangle of fixed edgelength e, and a procedure for making several "depth" measurements in the coordinate frame of the triangle, which is placed on the surface like a surveyor's tripod. These measurements take the form of

arc-lengths along "probe curves" at which the surface is intersected. Figure 1 shows three examples of TO's. Figure 1a shows a very simple TO with one line probe fixed symmetrically with respect to the rigid triangle ABC. The single scalar feature is the distance from the plane of ABC at which the probe intersects the surface. This resembles a mechanical optician's tool called a *spherometer*. We call the number d of scalar features the *order* of the operator. Figures 1b and 1c show TO's that can be viewed as a set of equilateral triangles hinged together so that all all d+3 points can be made to contact a surface. The angles of the d hinges are the features. We prefer this type (called *linkable* TO's) because of their symmetry and uniform sensitivity to noise. A planar surface yields $\theta = 180°$ for all d feature components. We will sometimes use $\phi \equiv \theta - 180°$ instead of θ for convenience. Many variations of these TO's could obviously be constructed. Feature noise is related to range noise n by the approximate expression $n_\phi \approx 51 \times n/e$, where n_ϕ is the feature error in degrees, and n is expressed in the same distance units as the edgelength e.
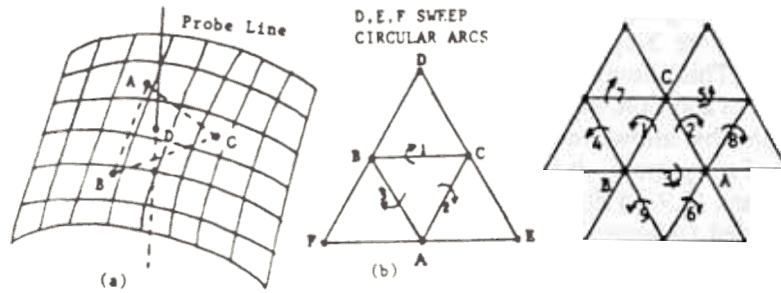

Figure 1. Examples of Tripod Operators: (a) Simple order 1 TO with linear probe, (b) Order 3 linkable TO, (c) Order 9 linkable TO.

For an N-point TO, the N sampled surface points yield only $N - 3$ independent scalar features (the *order* d is $N - 3$). These features contain all the surface shape information in the 3N components of the points, since they suffice to reconstruct the relative positions of the N points. They contain no other information; For example, they have complete six DOF invariance under rigid motions (the group $R^3 \times SO(3)$). Thus, they depend on where the tripod lies on the surface, but on nothing else. A key property is that only a 3-dimensional (at most) manifold of feature space points can be generated from a given surface, for any dimensionality d of feature vector, since the tripod can be moved only in 3 DOF on a surface. This allows objects to be densely sampled with TO's at preprocessing time with a manageable number of operator applications (typically, a few thousand) to obtain almost all the feature vector values obtainable from any range image of the object. This set is a kind of *invariant signature*. For brevity, we will call it the *signature* of the object or surface (with respect to a particular type of TO). It can be stored in bins (e.g., of dimension 3 or 4) for later efficient access of near neighbors to TO features measured at recognition time. These bins can optionally contain precomputed probability densities, analytic expressions for distances to nearby signature manifolds, and partial or complete descriptions of the relative poses of tripods and models, all to serve various purposes in a recognition system.

2.2 Computing a Tripod Operator Placement

Since in some applications of the tripod operator, the computation consists only of placement and a little indexing, the cost of placing the operator should be kept small. This can be done by efficiently implementing a procedure similar to the following. Consider placing the TO's of Figs. 1b or 1c on a dense range map. Point A can be chosen as any point on the image surface. Interpolation is to be done locally as needed (e.g., using piecewise triangular facets). Point B can be found by moving along a line at orientation α in image coordinates (pixel indices) until the 3D distance $|AB| \approx e$. This can be done in logarithmic time (essentially constant here) using binary search. Then we search the circle of radius $.5\sqrt{3}\,e$ oriented coaxially around the center of the segment AB, using binary search, to find a point C close to the surface. A similar circular search yields each remaining point. A key step in the circular search is the mapping (specific to a range scanner's geometry) from a point (x,y,z) to the indices of the range pixel

whose ray (x,y,z) lies on. This allows the front/behind decision required by the binary search. In the case of a sequential random access range scanner, it may be efficient to monotonically search elliptical paths in image coordinates until the two distances being enforced (e.g., |AC|=e and |BC|=e) are both correct. The ellipses here are the projections of the previously described circles onto image coordinates. Finally, in the case of a tactile TO, the computation is mechanical; the feature values are to be read from position transducers (e.g. from linear potentiometers by an A/D converter).

2.3 Symmetries of Surfaces and of Tripod Operators.

Surfaces with one symmetry, such as extrusions, surfaces of revolution, and helical projections produce only a 2-dimensional manifold in feature space. Cylinders, having two symmetries, produce only a nearly circular 1-dimensional curve, and spheres a single point. Scaling a TO by changing its edgelength does not effect the signature of surfaces swept by a line with one point fixed (e.g., cones, planar n-hedral vertices, and planar dihedral edges). Regardless of the surface, an operator with a 3-fold symmetry (e.g., those in Fig. 1), produces signatures unchanged by cyclicly permuting each triple of corresponding features. In Fig. 1c, the three 3-cycles (1,2,3), (4,5,6), and (7,8,9) show this property, for features ϕ_1 through ϕ_9, respectively. This allows a 3-fold storage reduction, e.g., by permuting the features so that ϕ_1 is the largest. If the TO, in addition, has handedness symmetry (as our examples do), the signature can be modified by a procedure that allows recognition of the "other side" of any surface already recognizable. We call this *inversion* of a signature. It is done by by transposing certain pairs of corresponding features (e.g., (7,5), (1,2), (4,8), and (6,9) in Fig. 1c) and replacing each feature value ϕ with $-\phi$. Also, the signature of the opposite-handed (reflected) version of a surface can be found by performing those transpositions without negating the features.

2.4 The Structure of the TO Signatures of Some Simple Shapes

We have been studying the shapes of TO signatures [17] in order to understand how they can overlap and to find ways to approximate them with algebraic and semi-algebraic expressions. Such approximations are expected to greatly reduce storage requirements for large libraries. The signatures of order 3 operators (Fig. 1b) were rendered as a rotating cloud of points on a computer; selected 2D snapshots are shown in Fig.2. In the special case of "smooth" surface regions, the signature is nearly a circular ring coaxial with the diagonal axis. The offset and radius of the ring can be readily used to compute estimates of the principle curvatures and other differential geometric parameters [17]. Surfaces with C_1 or C_2 discontinuities tend to produce signatures with similar numbers and kinds of discontinuities (e.g., Fig. 2c,d), and have roughly commensurate complexities of description. Thus, this umbrella-shaped 2-manifold can be well approximated with a few polynomials, whereas the discrete signature might need 20,000 points (see Fig. 4) for thorough saturation.
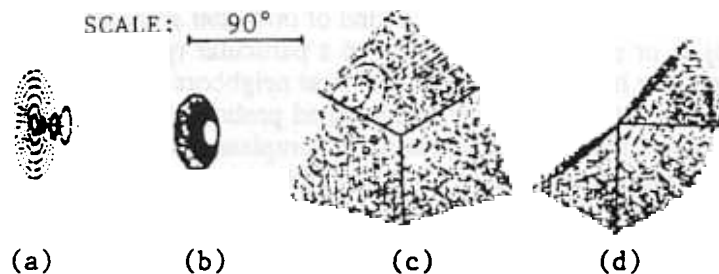
Figure 2. 2D projections of TO signatures taken with the TO of Fig. 1b. (a) Superimposed signatures of six hyperbolic paraboloid patches (large rings), four elliptic patches (rings lying on a cone), and 10 spheres (the points). (b) A torus; the signature is a piece of cone in $\phi_1\phi_2\phi_3$ space. (c,d) A 90° planar dihedral, viewed diagonally and along ϕ_1, respectively. All signatures of this TO have *at least* a 3-fold rotational symmetry about the diagonal $\phi_1=\phi_2=\phi_3$; all signatures in (a) and (b) are surfaces or curves of revolution.

3. CONDITIONS FOR RELIABLE RECOGNITION USING A SINGLE TO PLACEMENT

The low dimensionality of TO signatures (three, at most) typically allows the computation and storage of signatures containing (to a reasonable resolution) all feature vectors obtainable from a given surface shape, regardless of viewpoint. Moreover, since the feature space can have high dimensionality (d=9 in these experiments) the signatures of different objects' surfaces frequently have little or no intersection, allowing recognition of some objects with only one placement of a TO on the image of the object. Our experiments show that this circumstance occurs frequently with common shapes, and also that signature overlap can usually be dealt with. A deterministic viewpoint is taken here (When range error is a large fraction of TO edgelength, a probabilistic approach is essential).

3.1 False Positives

We will now derive sufficient conditions for precluding any false positive detections. Let us denote by A the set of all feature-space points obtainable by applying a certain TO to surface shape A. We call this the exact signature of shape A. Let A^δ denote some signature of shape A such that the greatest L_2 distance from any point in A to the nearest point in A^δ is δ. We call this a signature of A saturated to δ. This kind of signature can be obtained in practice by applying a TO a finite number of times to a surface. We similarly define B and B^δ for shape B. Now let A^- denote the set obtained by deleting from A^δ all points within an L_2 distance ε of any point in B^δ. We call this procedure *overlap removal* and speak of *subtracting* one signature from another. Now let v be the maximum L_2 distance that sensor (and other) error can introduce, and B^+ the set of points within v of B. Then B^+ includes all points actually obtainable by placing a TO on shape B. Summarizing key statements from above,

1. If $(b^+ \in B^+)$, $\exists (b \in B)$ s.t. $||b^+ - b|| < v$.
2. If $(b \in B)$, $\exists (b^\delta \in B^\delta)$ s.t. $||b - b^\delta|| < \delta$.
3. If $(a^- \in A^-)$ and $(b^\delta \in B^\delta)$, $||a^- - b^\delta|| > \varepsilon$.

Now consider a placement of a TO on shape B, producing the noise-corrupted feature point $b^+ \in B^+$ instead of the corresponding exact point $b \in B$. Suppose a^- is the nearest point in A^- to b^+. Then from 1., 2. and 3. above, $||b^+ - a^-|| \ge \varepsilon - \delta - v$. This means that we can never mistake a TO measurement taken from shape B for one taken from shape A using a threshold τ if $\varepsilon \ge \delta + v + \tau$. That is, if we index into the stored signature A^- using a measured feature value f, and find that $a^- \in A^-$ is within τ of f, then if $\varepsilon \ge \delta + v + \tau$, we are sure that f is not in the set B^+, and thus was not obtained from shape B. Figure 3 makes the inequality relation clear geometrically.
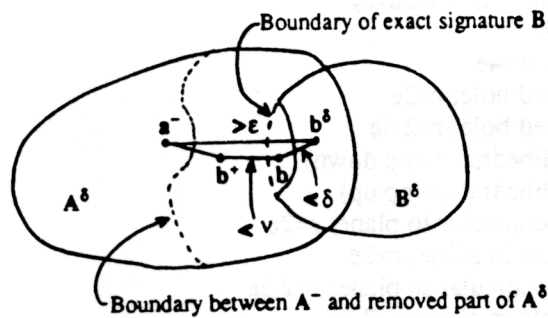
Figure 3. Schematic illustration of overlap removal for TO feature
space signatures

3.2 False Negatives

If $\tau < \delta + v$, note that a TO placement can fail to detect a shape due to insufficient saturation. This is generally of less importance than false positives, because negative classifications are simply the deferring of a decision, resulting in extra expended time to find an instance of the shape. If we want to be sure that

every TO measurement from shape A will lead to detection (allowing false positives from other shapes), we could test for nearness of the measured point to the signature A^+ using a threshold $\tau > \delta+v$. In section 4 we will see that the results of section 3 are overly stringent from a statistical point of view, e.g., we can violate $\varepsilon \geq \delta+v+\tau$ by a significant margin and still have very few false positives.

4. EXPERIMENTS

The purpose of these experiments is to study the discriminating power of an individual TO placement. Therefore, we use no preprocessing (except range rectification) and no hypothesis verification here. Nevertheless, this "pure" approach is quite powerful in many circumstances. In the experiments TO signatures were generated for 25 surface shapes. Next, overlap removal and analysis was done, followed by recognition experiments in which a specified shape is searched for until found.

4.1 Obtaining The Signatures

TO signatures were generated for 25 surface shapes by randomly placing an order 9 TO (Fig. 1c) on synthetic range images of each shape 50,000 times. The resulting signatures were stored as discrete feature-space points, with a numerical precision of $1°$. Duplicate feature vectors were removed, reducing the 50,000 points to as few as 61 points for the large cylinder and as many as 36,000 for the outside trihedral corner. Then the 3-fold symmetry of this TO (see section 2.3) was used to slightly increase the density of the signatures. These signatures correspond to A^δ of section 3, although δ was not directly controlled. The 25 shapes were chosen to include various discrimination challenges, e.g., cylinder vs torus with the same minor radius, and the hemisphere/cylinder (with C_2 discontinuity) dihedral region vs the cylinder or sphere. The following are the names and descriptions of the shapes ($e \equiv$ TO edgelength, $r \equiv$ radius):

0	plane	plane
1	cyl2e	cylinder; r = 2e
2	cyl2p5e	cylinder; r = 2.5e
3	sph2e	sphere; r = 2e
4	sph2p5e	sphere; radius = 2.5e
5	outcorner	outside 90° trihedral corner
6	ballcyl2e	hemisphere-capped cylinder; r=2e
7	ballcyl2p5e	hemisphere-capped cylinder; r=2.5e
8	incorner	inside 90° trihedral corner
9	pcyl2e	plane-capped cylinder; r=2e
10	pcyl2p5e	plane-capped cylinder; r=2.5e
11	dh270	270° planar dihedral (convex)
12	dh90	90° planar dihedral (concave)
13	tor2e4e	torus; r=2e, R=4e
14	tor2p5e4e	torus; r=2.5e, R=4e
15	phole2e	plane-bottomed hole; r=2e
16	phole2p5e	plane-bottomed hole; r=2.5e
17	dh225	225° planar dihedral (ramp down)
18	dh135	135° planar dihedral (ramp up)
19	peg2e	cylinder perpendicular to plane; r=2e
20	edgehole2e	cylindrical hole in plane; r=2e
21	peg2p5e	cylinder perpendicular to plane; r=2.5e
22	edgehole2p5e	cylindrical hole in plane; r=2.5e
23	thshelf	planar trihedral; 90°, 90°, 270°
24	thnotch	planar trihedral; 90°, 270°, 270°

Some are inversions of each other; (5,8), (23,24), (9,15), (10,16), (11,12), and (17,18). In these cases we generated the latter by inverting the data from the former (see section 2.3). We see in Table 1 that most pairs of the 25 shapes' initial signatures were already entirely disjoint (separation > 5°) including a cylinder (2) and the torus (14) with the same minor radius. Most ambiguous points were from shared

parts; an inside trihedral corner (8) contains an inside dihedral edge (12). Later, we will use overlap removal to make the final signatures (nearly) disjoint by design.
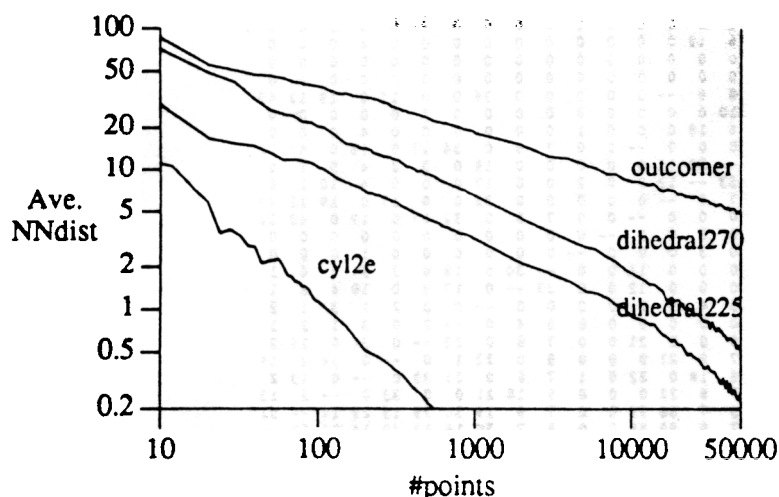

Figure 4. Saturation graphs for four representative shapes.

The results of section 3 show the importance of highly *saturated* signatures (small δ). Therefore we have studied the dependence of the degree of saturation on the number n of randomly placed order 9 TO placements, for various shapes. Figure 4 shows log/log plot of the average L_2 distance σ (in degrees) of a feature space point to its nearest neighbor versus n. We found that the dependence is approximately $\sigma = c/n^{1/k}$, where k is approximately the dimension of the signature manifold. k = .952 for the cylinder, whose manifold has dimension 1. k = 1.89 and 1.92 for the 225° and 270° dihedrals, respectively, whose manifolds have dimension 2, and k = 2.94 for the fully three dimensional outside corner. The k values are slightly lower than the corresponding dimensionality primarily because of low-dimensional subshapes (e.g., the plane (k=0) is in the n-hedral shapes). These empirical results are consistent with the observation that the density of n random points on a k-manifold is approximately proportional to n^k. Note that σ is not the same as δ; e.g., for pcyl2p5e, $\sigma=3.4$, while about .1% of new points from this shape were farther than 10° from their nearest neighbor in our signature. Thus $\delta > 10°$.

4.2 Signature Overlap
Next, pairs of signatures were processed to remove overlap ($\epsilon = 5°$) with other shapes' signatures. Certain of these "set subtractions" were forbidden; e.g., we did not allow shapes that are parts of other shapes to be deleted. For example we did not "subtract" dh90 from plane. The full set of forbidden pairs is (5,8,9,10,11,12,15,16,17,18,19,20,21,22,23,24) from 0, (6,9,19) from 1, (7,9,10,21) from 2, 6 from 3, 7 from 4, (5,23,24) from 11, and (8,23,24) from 12, referring to the list above. Table 1 was computed before overlap removal, showing the percentage of shape A left after subtracting shape B, for all 25^2 pairs. Note that most of the signature pairs have little or no overlap, allowing easy discrimination.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
0:	--	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1:	0	--	0	0	0	0	25	0	0	9	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
2:	0	0	--	0	0	0	0	28	0	6	18	0	0	0	0	0	0	0	0	0	0	4	0	0	0
3:	0	0	0	--	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4:	0	0	0	0	--	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5:	--	0	0	0	0	--	0	0	0	8	9	--	0	0	0	0	36	0	0	17	0	18	12	43	
6:	0	--	0	--	0	0	--	0	0	10	0	0	0	0	0	0	0	0	0	3	0	0	0	0	
7:	0	0	--	0	--	0	0	--	0	6	18	0	0	0	1	0	0	0	0	0	0	4	0	0	
8:	--	0	0	0	0	0	0	0	--	0	0	0	--	0	0	7	9	0	36	17	0	18	0	43	12
9:	--	--	--	0	0	7	26	26	0	--	30	11	0	0	0	0	0	18	0	3	8	4	9	1	6
10:	--	0	--	0	0	8	0	31	0	33	--	12	0	0	2	0	0	17	0	9	4	10	1	6	
11:	--	0	0	0	49	0	0	0	7	9	--	0	0	0	0	0	32	0	0	17	0	19	12	43	
12:	--	0	0	0	0	0	0	0	49	0	0	0	--	0	0	7	9	0	32	17	0	19	0	43	12
13:	0	0	0	0	0	0	0	0	0	0	0	0	0	--	0	0	0	0	0	0	0	0	0	0	
14:	0	0	16	0	0	0	0	3	0	0	3	0	0	0	--	0	0	0	0	0	0	1	0	0	0
15:	--	0	0	0	0	0	0	0	7	0	0	11	0	0	--	30	0	18	8	3	9	4	6	1	
16:	--	0	0	0	0	0	0	0	8	0	0	0	12	0	0	33	--	0	17	9	0	10	4	6	1
17:	--	4	10	0	0	3	2	2	0	5	4	5	0	0	0	0	0	--	0	0	2	1	3	1	2
18:	--	0	0	0	0	0	0	0	3	0	0	5	0	0	5	4	0	--	2	0	3	1	2	1	
19:	--	--	0	0	0	0	26	0	13	9	0	0	21	0	0	7	8	0	23	--	0	34	0	14	2
20:	--	0	0	0	0	13	0	0	0	7	8	21	0	0	0	9	0	23	1	0	--	0	34	2	14
21:	--	0	--	0	0	0	0	28	13	6	18	0	22	0	1	7	8	0	21	33	0	--	0	13	2
22:	--	0	0	0	0	13	0	0	0	7	8	22	0	0	6	18	21	0	0	33	0	--	2	13	
23:	--	0	0	0	0	43	0	0	49	5	7	88	99	0	0	7	9	19	30	29	13	29	14	--	51
24:	--	0	0	0	0	49	0	0	43	7	9	99	88	0	0	5	7	30	18	13	29	14	29	51	--

Table 1 Overlap percentages (-- denotes 100%) for raw signatures, before overlap removal; shape indexed at left *subracted* from shape indexed at top, with separation threshold ε=5°.

4.3 Recognition

Each signature was stored in bins in a three-dimensional array (using the first 3 feature components) to facilitate near-neighbor lookup. At recognition time our system randomly placed TO's on the synthetic range image of Fig. 5a, which contains instances of all 25 shapes, and labeled the locations of the TO's as ambiguous or unknown (white) or as the shape currently being sought (black). The decision rule was to note whether the distance from the TO feature vector to the nearest point in the signature at hand was less than τ, which was set to 5°. For each of the 25 shapes, with range noise initially zero to help isolate error sources, we applied the TO enough times to obtain 50 correct detections of the shape, and recorded various results such as the mean time (in TO operations) between detections (MTBD) and data on any false positive detections. The MTBD can be regarded as the ratio of the image area to the "effective area" of the shape sought, for a particular TO size e.
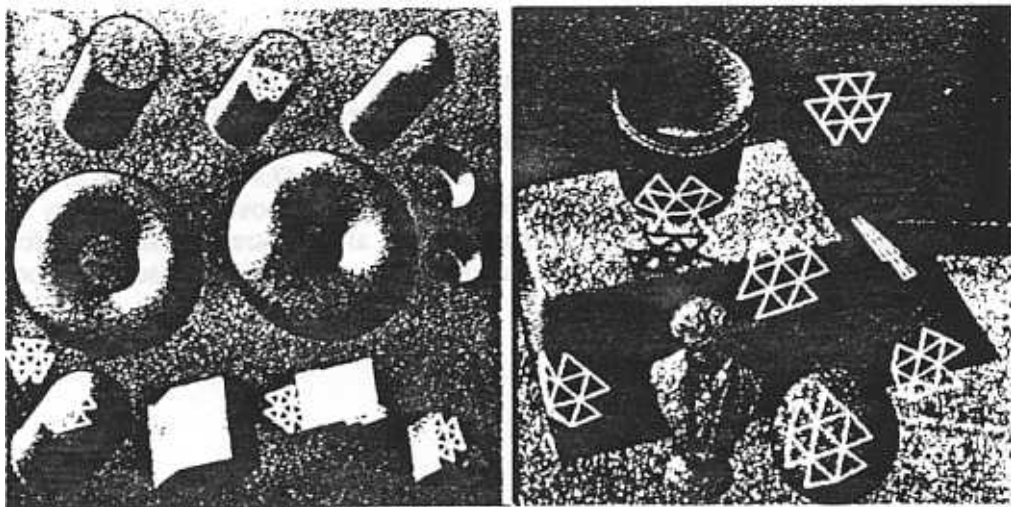
Figure 5. Noisy range images showing reliable detection of shapes by TO's; (a) tor2p5e4e is detected in 8 placements in a synthetic image. (b) dh270 is detected in 8 placements on a LIDAR image with TO edgelength = 7 cm. Both took approximately 50 milliseconds.

About 63% of the TO placements on the image of Fig. 5a *aborted* due to contact of a probe point with a jump boundary, which is locally detected by pixel disparity. This is typical for cluttered scenes and is a highly efficient substitute for image segmentation. The following results are described for non-aborted placements. The smallest MTBD values were for plane (2.74) and the tori (both about 5). The largest was for phole2e (483), the small plane/cylinder dihedral at the bottom of the small hole.

The estimated mean time between false positive detections (MTBF) was ∞ (none observed in several thousand placements) for 12 shapes and varied from 17,088 placements for incomer to 127 for pcyl2e (pcyl2p5e was falsely detected). Overall, the results showed very few false positives, which were primarily due to the lack of sufficiently exhaustive signatures, leading to failure to delete some point common to two shapes. False positives due to unknown objects are more difficult to prevent, but are fairly rare. The false positives are all due to violation of $\epsilon \geq \delta+v+\tau$. If δ were 0 (exact signature), $5° \geq 0+0+5°$ would hold, precluding false positives. However, δ exceeds $10°$ in some of the signatures used, due to small portions of the exact signature B being farther than $10°$ from the stored signature. This causes no trouble for most pairs, because they are already separated by much more than the imposed ϵ before overlap removal. However, our pcyl2p5e signature has both high δ and high overlap with pcyl2e, causing the above problem. We found that we could drive sharply down the false positive incidence by either increasing ϵ, which had the side effect of increasing the MTBD by introducing false negatives, or by sampling more to decrease δ, particularly at the low density places in the signature. The latter is more attractive, because it does not compromise the MTBD rates significantly. We plan to pursue the construction of uniform density signatures with tightly controlled δ to address this issue.

Having discussed how to avoid shape confusions in the absence of noise, we ran some recognition examples with added range noise of peak value e/40 (edgelength e=.2, noise = .005). This yielded a peak displacement $v=6.5°$ in feature space. For example, for the two tori, with $\tau = 5$, we found that at this noise level, well within the capabilities of various existing range sensors [15,17], there were no false positives in thousands of trials, and only 10% of the TO's falling on the tori failed to detect them. Their signatures are only $7°$ apart at their nearpoints. This violates $\epsilon \geq \delta+v+\tau$ ($\epsilon=7$, $\delta=1$, $v=6.5$, $\tau=5$), but the probability of the vectors in Fig. 3 aligning just right to cause a false positive appears small, both considering the geometry of Fig. 3 and the experiments. Repeating this with range noise e/20, we found that the large torus was mistaken for the large cylinder 10% of the time (*all* of cyl2p5e lies within $6.8°$ of tor2p5e4e), but all tor2e4e detections were correct. In figure 5b, we search a LIDAR image for the dihedral dh270 with range noise $\approx$ e/23 (~3mm). The MTBD is about 15 placements, and we saw no false positives. Our next step will be to seek a systematic way to set ϵ, τ and δ, given the average noise, for optimal performance.

5. CONCLUSIONS AND FUTURE DIRECTIONS

We have studied the ability of individual order 9 TO's to discriminate 25 surface shapes in a cluttered range image and concluded they can in many circumstances do so rapidly and with very few false positive detections. Conditions for guaranteeing this were derived. A TO can be applied and interpreted in less than 2 milliseconds on a Sparc Workstation. We plan to reduce this time by software optimization, and to study analytic approximations of TO signatures, combining pose constraints using multiple TO placements, probabilistic approaches, and other topics aimed at finding the limits of their performance.

We believe that one of the most promising potential applications of TO's is the recognition and localization of industrial parts, because (a) TO's provide the speed frequently required for economic viability, (b) range scanner technology has been rapidly achieving the requisite resolution, speed and cost, and (c) the market is broad, including automatic assembly, inspection and materials handling. Other potential applications include automatic target recognition with LADAR, a portable "seeing eye stick" for the blind that names obstacles such as "step down", pole $20°$ left, etc., and AGV navigation by recognizing existing indoor features. The latter might be worthwhile if the same vision system is used to recognize objects in "gopher" tasks. Also, we plan to study the effectiveness of TO's to face recognition, using a statistical approach similar to that in [3]. In general, TO's are applicable to the recognition of a wide variety of

objects for which reasonably accurate range images can be economically obtained. Exceptions include thin wire-like objects, on which the sample points of the TO are unlikely to fit, and objects for which the reflectance markings contain essential information. Some of the latter may yield to a registered range and intensity approach that unifies tripod operators with existing OCR methods.

REFERENCES

[1]   Pipitone, F., "Tripod Operators for the Interpretation of Range Images", Naval Research Laboratory Memorandum Report #6780, February, 1991.

[2]   Pipitone, F., and Adams, W., "Tripod Operators for Recognizing Objects in Range Images; Rapid Rejection of Library Objects", Proc. IEEE International Conf. on Robotics and Automation, pp1596-1601, Nice, France, May 1992.

[3]   Pipitone, F., and Adams, W., "Rapid Recognition of Freeform Objects in Noisy Range Images Using Tripod Operators", Proc. IEEE Conf. on Computer Vision and Pattern Recognition, New York, NY, June, 1993.

[4]   Grimson, W.E.L., and Lozano-Perez, T., "Model-Based Recognition and Localization from Sparse Range or Tactile Data", The International Journal of Robotics Research, Vol. 3, No. 3, pp 3-35, Fall 1984.

[5]   Grimson, W. E. L., "The Combinatorics of Object Recognition in Cluttered Environments Using Constrained Search", MIT AI Memo No. 1019, February, 1988.

[6]   Gaston, P. C., and Lozano-Perez, T., "Tactile Recognition and Localization Using Object Models: The Case of Polyhedra On A Plane", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6 (3):257-265, May 1984.

[7]   Oshima, M. and Shirai, Y., "Object Recognition Using Three-Dimensional Information", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-5(4):353-361, July, 1983.

[8]   Faugeras, O.D., and Hebert, M., "A 3-D Recognition and Positioning Algorithm Using Geometrical Matching Between Primitive Surfaces", Proc. Eighth Int. Joint Conf. Artificial Intelligence, pp 996-1002, August, 1983.

[9]   Bolles, R.C., and Cain, R.A., "Recognizing and Locating Partially Visible Objects: The Local-Feature-Focus Method", International Journal of Robotics Research 1(3):57-82, 1982.

[10]   Lamdan, Y., and Wolfson, H.J., "Geometric Hashing: A General and Efficient Model-Based Recognition Scheme", IEEE 2nd International Conf. on Computer Vision, 1988.

[11]   Stein, F., and Medioni, Gerard, "Structural Indexing: Efficient 3-D Object Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-14 (2):125-145, February, 1992.

[12]   Fischler, M.A., and Bolles, R.C., "Random Sample Consensus, a Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", Comm. of the ACM, v24,n6, pp381-395.

[13]   Besl, P.J., and Jain, R.C., "Invariant Surface Characteristics for 3D Object Recognition in Range Images", Computer Vision, Graphics, and Image Processing 33, 33-80, 1986.

[14]   Bolles, R.C., and Horaud, P., "3DPO: A Three-Dimensional Part Orientation System", International Journal of Robotics Research, 5(3): 3-26, 1986.

[15]   Rioux, M., Blais, F., Beraldin, J., and Boulanger, P., "Range Imaging Sensors Development at NRC Laboratories", Proc. of the Workshop on Interpretation of 3-D Scenes, pp 154-160, Nov. 27, 1989, Austin, TX, IEEE Press.

[16]   Besl, P.,"Active, Optical Range Imaging Sensors", in *Machine Vision and Applications*, Springer-Verlag, Vol. 1, pp 127-152, 1988.

[17]   Pipitone, F., "Extracting Elementary Surface Features Using tripod Operators", Internal Report AIC-93-036, Navy Center for Artificial intelligence, US Naval Research Lab, Washington, DC, 1993.